Bayesian Analysis (2007) **2**, Number 1, pp. 31–36

Comment on Article by Dominici et al.

Samantha R. Cook* and Elizabeth A. Stuart†

1 Introduction

Causal inference, and specifically principal stratification (Frangakis and Rubin 2002), is an important area of statistics with the potential to answer fundamental questions in medicine, economics, and many other scientific disciplines. Dominici et al. have done a commendable job of applying and extending the ideas of principal stratification to address specific questions about the impact of vitamin supplementation on birth weight and infant mortality. We applaud and congratulate the authors on an insightful and important paper. After briefly reviewing the paper, we focus on three issues: the assumptions required for principal stratification applications such as this one, the particular causal quantities considered here, and possible model extensions to handle observational data or more complex outcomes.

This paper considers the effect of vitamin supplementation on infant mortality; analysis is aided by the fact that treatments were assigned randomly and most study participants complied with their assigned treatment. Inference was complicated, however, because the treatment effect was believed to be non-constant. Estimating causal effects conditional on covariates is straightforward; in this case, however, the causal effects of vitamin supplementation on mortality were believed to vary with birth weight, which is itself an outcome that may depend on the treatment received. We therefore have a primary outcome, infant mortality, that may be related to an intermediate outcome, birth weight. The authors focus on two quantities of interest: the percentile-specific effects of supplementation on birth weight and the effects of supplementation on infant mortality, principally stratified by birth weight.

2 Principal Stratification and Associated Assumptions

A causal effect is fundamentally a comparison of two *potential outcomes*: the outcome a single individual would experience if assigned to take the treatment and the outcome the individual would experience if assigned to control (Rubin 1977). Because we can observe at most one potential outcome on each unit, causal inference is inherently a missing data problem (Holland 1986). Inference tends to focus on average treatment effects or average treatment effects within subgroups rather than individual causal effects, which are never observed. When subgroups are defined by pre-treatment covariates and treatments are assigned randomly, causal inference remains relatively straightforward. Estimating treatment effects conditional on *post-treatment* variables is more complicated, however, because individuals with similar values of post-treatment variables are not necessarily

*Columbia University, New York, NY, mailto:cook@stat.columbia.edu
†Mathematica Policy Research, Inc, Princeton, NJ

comparable in the way that individuals with similar values of pre-treatment variables are comparable, even if treatments are assigned randomly. As an example, consider the subgroup of infants who are normal (not low) birth weight. Among these infants, those in the control group are likely to be healthier on average than those in the treatment group. Assuming that supplementation affects birth weight, some of the normal birth weight infants who did receive supplements would likely have been low birth weight if they had not received supplements. Stratifying the infants based only on observed birth weight does not utilize the random assignment of infants to treatment and control groups and can therefore result in comparisons between infants who are fundamentally different from one another; these comparisons will not in general produce valid causal estimates. Stated another way, stratifying on observed birth weight forms subgroups using only the observed values of what is actually a bivariate outcome consisting of both potential outcomes. On the other hand, the *vector* of *potential* post-treatment variables is not affected by treatment assignment; estimates calculated within strata defined by this vector can still represent valid causal quantities. This idea, known as principal stratification, underlies some forms of economic instrumental variables estimation and was formalized by Frangakis and Rubin (2002).

Because the vector of potential post-treatment variables is never fully observed for any individual, principal stratification estimation relies on assumptions about the missing potential post-treatment variables, that is, the post-treatment variables that would have been observed under the treatment not received. Most work to date in principal stratification (e.g., Barnard et al. 2003; Zhang and Rubin 2003) has focused on binary post-treatment variables, where certain assumptions can identify missing potential outcomes. Consider the simplified situation where birth weight takes on one of two values: high or low. In this case, assumptions that might be considered are monotonicity (there are no infants who are low birth weight with supplementation and high birth weight without supplementation) or exclusion restrictions (for those infants whose birth weight is not affected by supplementation, mortality is also not affected by supplementation).

With continuous post-treatment variables, such as birth weight, more complicated models with different kinds of assumptions are needed. In particular, the causal quantities considered by Dominici et al. will depend on the correlation between infants' two potential birth weight outcomes (birth weight with supplementation and birth weight without supplementation). Because these two potential outcomes are never observed for the same infant there is no direct information about this correlation in the data. Therefore, any inferences will rely on assumptions about this correlation. Dominici et al. fix the unknown correlations between potential outcomes (both birth weight and mortality) at plausible values, and then vary these values in a sensitivity analysis. As another type of sensitivity analysis, these unknown correlations could be modeled using prior distributions that are uniform over the range of values considered in the sensitivity analysis, thereby producing estimates that are smoothly averaged over the uncertainty about the correlation parameters. Assuming fixed values of these correlations, as Dominici et al. do, is simply an alternate (point mass) prior specification; we feel that estimates based on both types of prior distribution could provide complementary information. This is an example where substantive prior knowledge about the possible magnitudes

of these correlations is imperative, and underscores the importance of collaboration and communication between statisticians and subject-matter experts.

3 Quantities of Interest

The authors present two types of percentile-specific treatment effects of supplementation on birth weight: a "population" estimate based on the percentiles of the marginal distributions of $W(0)$ and $W(1)$ and a "causal" estimate based on the distribution of $W(1) - W(0)$, where $W(Z)$ is the observed birth weight when assigned to treatment Z. These two quantities represent a trade-off between what is easily estimable and what is scientifically interesting. The population estimates are interpretable as causal only if $W(1)$ is a monotone function of $W(0)$. This ensures that the relative ordering of the infants is the same when sorted by either $W(0)$ or $W(1)$ and thus ensures that the causal effect estimates are comparisons of potential outcomes on common sets of infants. The causal estimates, however, rely on assumptions about ρ, the unestimable correlation between $W(1)$ and $W(0)$, whereas the population estimates do not depend on ρ. As ρ approaches 1, the two estimates should converge. The distinction is perhaps less important, then, in cases such as this one where the correlation between potential outcomes is believed to be fairly large.

Of primary interest are the principal stratification estimates of the causal effect of supplementation on infant mortality. The causal effects are stratified by two levels of birth weight under control (less than or greater than 2500 g) and two levels of the causal effect of supplementation on birth weight (absolute value of the causal effect less than or greater than 50 grams). The results are intuitively appealing: The subgroup for which there appears to be a significant positive effect of supplementation on mortality consists of those infants who would be low birth weight without supplementation and whose birth weight changes by 50 g or more with supplementation. These results help clarify the way in which supplementation can decrease infant mortality and shed light on the scientific mechanisms underlying the observed data. It strikes us as odd, however, that the principal strata are defined by the absolute value of the causal effect of supplementation on birth weight. Infants who would be low birth weight without supplementation and who would have even lower birth weight with supplementation seem to represent a fundamentally different group from those who would be low birth weight without supplementation and who would have higher birth weight with supplementation. We might expect the causal effects of supplementation on mortality for these two sub-subgroups to differ substantially, obscuring the effect estimates. We acknowledge that this may merely be a notational issue: Presumably the proportion of this subgroup whose birth weight decreases rather than increases with supplementation is small and perhaps negligible. An assumption stating this formally would be a continuous analogue of the monotonicity assumption described earlier: Birth weight with supplementation cannot be lower than birth weight without supplementation. To the extent that this group exists, however, we think a more useful way to stratify the causal effects is by high/low birth weight under control and high/low *increase* (rather than change) in birth weight.

For clinical practice, it would also be interesting to know what proportion of the population falls into each of the four principal strata and whether there are any observable covariates that predict well the infants' stratum membership. A difficulty in defining the subgroups by the causal effect of supplementation on birth weight is that this quantity is unobservable for each infant. Given the results found by Dominici et al., which indicate varying impacts for infants in different principal strata (including potentially negative impacts for some infants), it would be particularly useful to be able to predict the relative efficacy of the supplements for a particular infant.

4 Extensions

We expect that the methods presented here will be useful in many other applications; we elaborate on one and discuss possible extensions. A question of interest to economists and criminologists is the effect of post-prison reentry programs on recidivism for parolees. Such programs often involve job training, and an important post-treatment variable in this case is income: Parolees who go on to earn a living wage after being released from prison might be expected to have lower rates of recidivism than those parolees who earn low wages or are unemployed. (An example of such a program is Exodus Transitional Community, Inc., in East Harlem, NY, operated in conjunction with Ready4Work, a government-funded organization. See *http://www.dol.gov/cfbci/Ready4Work.htm.*)

Such a situation presents at least two additional complications beyond those addressed in this paper. First, wages are generally considered undefined (as opposed to taking the value zero) for people who are unemployed. This means that we really have two post-treatment income variables: an employment indicator, and wages for those who are employed. One possible poststratification of interest in this situation consists of the following subgroups: (1) those who are unemployed under both treatments, (2) those who are employed under treatment but unemployed under control, (3) those who are unemployed under treatment but employed under control, (4) those who are employed under both treatment and control and whose wages do not increase significantly with treatment, and (5) those who are employed under both treatment and control and whose wages do increase significantly with treatment. Groups 2 and 5 would be of primary interest for understanding the relationship between employment and recidivism. Such an analysis would require even more structure and assumptions about both employment and wages in order to identify causal effects. For example, the correlation between wages under the two treatments is only defined for the subgroup who would be employed under both treatments. This situation would be somewhat analogous to that presented in Dominici et al. if their analysis were extended to consider all pregnancies rather than only live births, since birth weight is defined only for live births. Limiting the analysis to only live births does represent selection based on an outcome and, to the extent that supplementation affects miscarriages, could obscure inferences.

Another complication is that such reentry programs tend to be voluntary, meaning that the backdrop of a randomized experiment no longer exists. Those people who

choose to participate in a reentry program might be less likely to return to prison regardless of their actual participation in such a program. Even if such a program were to be randomly assigned to participants, we expect that compliance to treatment assignment would be low enough that additional assumptions or methods for observational studies would be needed. One possibility is using propensity score methods to choose a subset of people who are similar on relevant background characteristics (education, time spent in prison, social support, etc.), where some received the treatment and others did not. If treatments were assigned randomly, another option is to consider treatment compliance as an additional post-treatment variable. The principal strata could then depend on treatment compliance, employment, and wages.

5 Conclusion

The authors have made an important step in the estimation of causal effects, and we applaud them on a superb paper. In addition to its value for causal inference, this paper also highlights some of the strengths of Bayesian analysis. Treating all unknowns as random quantities allows for prediction of missing potential outcomes and of the underlying values of birth weight when measured with error. In addition, assumptions about unknown quantities such as the correlation parameters ρ and ψ are made explicit and different prior distributions on these quantities can allow for different types of sensitivity analyses. Incorporating covariates into the analysis is also straightforward, as well as critical in these types of problems.

References

Barnard, J., Frangakis, C., Hill, J., and Rubin, D. (2003). "Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City (with discussion)." *Journal of the American Statistical Association*, 98: 299–323. 32

Frangakis, C. and Rubin, D. (2002). "Principal stratification in causal inference." *Biometrics*, 58: 21–29. 31, 32

Holland, P. W. (1986). "Statistics and Causal Inference." *Journal of the American Statistical Association*, 81: 945–60. 31

Rubin, D. B. (1977). "Assignment to treatment group on the basis of a covariate." *Journal of Educational Statistics*, 2: 1–26. 31

Zhang, J. and Rubin, D. (2003). "Estimation of causal effects via principal stratification when some outcomes are truncated by "death"." *Journal of Educational and Behavioral Statistics*, 28: 353–368. 32